

Feature Selection using Multi-Objective Clustering based Gray Wolf Optimization for Big Data Analytics

Aakriti Shukla
Department of CSE
Bansal Institute of Science & Technology
Bhopal (M.P.), India
aakritishukla2512@gmail.com

Dr. Damodar Prasad Tiwari
Department of CSE
Bansal Institute of Science & Technology
Bhopal (M.P.), India

Abstract: Although numerous efforts have been made to develop feature selection framework which is efficient in Big Data technology, complexity of processing big data remains a significant barrier. As a result, the computational complexity and intricacy of big data may block the data mining process. The feature selection method means, a required pre-processing approach to minimize dataset dimensionality for great advanced features and classifier performance optimization. In order to increase performance, feature selection are regarded to constitute the core of big data technologies. In recent years, many academics have moved their focus to data science and analytics for application scenarios leveraging integrating tools of big data. People take quite some time to engage, when it comes to big data. As a consequence, in a decentralized system with a high workload, it is crucial in making feature selection dynamic and adaptable. Multi objective optimal strategies for feature selection are provided in this work. This research adds to the creation of a strategy for enhancing feature selection efficiency in large, complex data sets. In this paper, a multi-objective clustering-based gray-wolf optimization algorithm (MOCGWO) is proposed for classification problems. Five datasets were used to show the robustness of proposed algorithm. The result analysis was compared with other optimization methodology such as GWO and PSO. This shows efficacy of MOCGWO algorithm.

Keywords: Feature Selection, Big Data, Classification, Multi-objective, Clustering, Optimization.

I. INTRODUCTION

Artificial Intelligence (AI) approaches have recently gotten a lot of interest in a lot of applications because of their capacity to extract unexpected data. In view of the massive amounts of data collected each day and the necessity to interpret this data into useful information, Big Data approaches are found in many of the most prominent data processing techniques, such as data mining, ML, AI, data fusion, online network site etc[1]. Big Data Analytics is recognized as a significant AI tool. Knowledge discovery with Big Data are at the frontline of several technologies in the sectors of telecommunications,

medical data analysis, account records, and so on. The main objective of Big Data Analytics is to acquire and use of knowledge which flows as a result of dealing with massive data sets. Big Data is described as a collection of data with a size, velocity, or diversity so big that typical databases as well as data processing systems struggle to save, organize, process, but also analyze the data. Big Data Analytics are techniques and tools used to analyze and extract information from Big Data [2].The process of deriving insight from raw data by studying and analysing its patterns and attitudes using qualitative and quantitative methodologies is referred to as Big Data analytics. Using it in healthcare offers enormous promise for increasing patient care quality, decreasing waste and error, and lowering costs.[3] Advanced analytics may help you make better decisions, reduce risks, and get useful insights from your data. As a result, big data methods and technology must be used to supplement judgments by examining large databases.

Traditional methods cannot examine such a massive number of data; therefore, sophisticated techniques known as big data analytics must be used.

Advanced tools developed to deal with massive amounts of heterogeneous data in order to enhance the conventional Data Analytics Process are referred to as "big data analytics." Big Data Analytics has grown in popularity in recent years and is now being utilized in a variety of settings including healthcare, government administration, finance, management, and telecommunications. Amongst the most essential jobs in ML is FS. The Purpose of feature selection is to chose a limited number of variables while yet attaining good classification results. It is possible to reduce the complication of the resulting learning model as well as simplifying the interpretation by selecting a subset of features also with fewest inner similarity and even the greatest relevance to the target class[4]. Many characteristics are present in physical ML tasks, but not all of them are required for categorization. Here, 2 kinds of approaches of feature selection: filter-based, and wrapper-based. Data-reliant requirements are used in the filter-based procedure for analyzing qualities of the feature subset. Meanwhile wrapper-based technique, used a specific classifier to evaluate the

important features [5]. The characteristics that are irrelevant or duplicated may have a detrimental influence on categorization. To begin with, they greatly slow down learning. Furthermore, they distort the learning model by include characteristics that are extraneous to the learning process. Lastly, they degrade classifier performance. To increase learning efficiency as well as results in terms of accuracy, it is crucial to eliminate unnecessary and redundant information. The basic goal of the FS issue always to lower feature set's size while preserving classification performance [6]

The multiobjective optimization problem (as well known the multiobjective programming issue) is an area of mathematics that specializes with optimization problems containing 2 or more objective functions that must be optimized at the same time. Whether supervised Feature selections or unsupervised Feature selections problems, both expressed like a single objective (SOO) or a multi objective (MOO).[7]. Whether the classification strategy is supervised or unsupervised, implementing FS like a multi-objective optimization process may provide certain advantages. Because the 2 objectives often in conflict, here FS is described by multi-objective optimization method.[8] A possible description of this issue is a multi-objective optimization method that adequately blends classifier performance with attribute amount. Multi-objective optimization refers to problems having several objectives (MOO). The goal of this study is to develop Multi-Objective Optimization Based FS techniques' for large data feature selection.

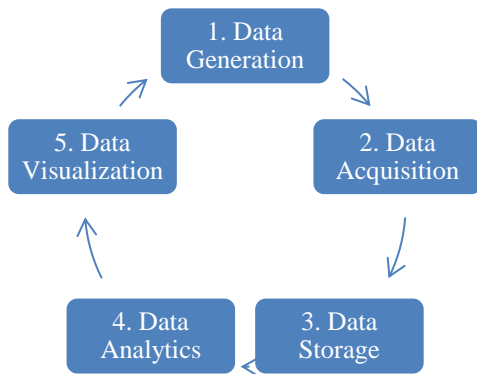


Fig. 1. Big Data Analytics Flow

The MMO algorithms are utilized to address feature selection issues in this research. Several multimodal optimization techniques have been suggested[9]. It can identify numerous feature subsets along with the amount of features as well as classification error rate, however different feature combinations. It will give consumers with not only the quantity of features, but also the locations of the features they have chosen. The users may then make a judgment in big data based on their preferences.

II. RELATED STUDY

In [10] author considered feature selection (FS) and classifying problems as critical issue. Its goal is to find the traits that are both meaningful and non-redundant. Given the large amount of features in practical uses, FS approaches based on batch learning techniques are unable to tackle big data challenges, particularly when data arrives in a proper sequence. In this research presenting an online FS technique to overcome this issue. The suggested Online Feature Selection system, Multi-Objective Automated Negotiation based Online Feature Selection (called MOANOFS), aims to improve the classification accuracy of mega dimensional databases by combining recent advances in online ML techniques with a conflict resolution technique. MOANOFS employs two stages of decision making. At the first, determining which of the learners' k(s) (or OFS approaches) are the most reliable (with trust value). Then k learners move to the second level, implementing suggested Automated Negotiation called Multilateral Automated Negotiation based OFS (MANOFS) technique. It could assist in determining the key traits. With many real-text categorization datasets, like 20Newsgroups and Reuters Corpus Volume I, we demonstrated that the MOANOFS system achieves excellent accuracy.

In [11] suggested SMTIBEA a hybrid multi-objective programming technique which merges Indicator-Based Evolutionary Algorithm (IBEA) plus Satisfiability Modulo Theories (SMT) solutions. SMTIBEA greatly improves the flexibility of constraint expressions in practically SPLs by extending the state of the art beyond conjunctive normal form (CNF) Boolean formulae upto involving first order quantifier free formulas. Testing the techniques on 5 big, physical, highly limited SPLs and assessed them using five performance measures. Given that the expressiveness of restrictions has been greatly increased, empirical data show that our technique is equivalent in efficiency to the state-of-the-art. To study the performance impact of SMT solving upon that mutation operator as well as the initialization population generation of the IBEA, creating three SMTIBEA versions. The value of combining Mutation or genetic operator by constraint solving such as constraint satisfaction was confirmed by our empirical findings. It is advised to supplement the initial population generation in Satisfiability Modulo Theories solving to obtain solutions of such higher HV. The expressiveness of technique on more difficult SPL configuration optimization challenges will be investigated in future work.

In [12] author proposed big Data optimization to optimization challenges that involve big data analytics features' management. Search Manager (SM), a newly suggested techniques for combining different algorithms hybrid of metaheuristics to enhance productivity of methodology techniques, get stretched for (MOSM) in this article, then 5 configurations itself are suggested using a mixture of multiple search strategies to overcome the issues regarding electroencephalogram signal analysis, that became

part of the BD optimization issues class. The suggested MOSM configurations were effective in such situation, according to experimental data. The assortment are contrasted here to Non-dominated Sorting Genetic Algorithm NSGA-III UCAM, a newly suggested Big-Opt approach.

In [13] provided the high computing costs of classic multi-objective evolutionary Algorithms (MOEAs) that significantly difficulty for solving (Big-Opt) challenges. The 3rd update of the NSGA-III, has successfully addressed this problem. Furthermore, another of the NSGA-III individual's flaws is employing a rigid price for the mutation operator. For confronting that, developing an mutation operator that aims for improving traditional NSGA-III algorithm's efficiency. 3 NSGA-III crossover operators, simulated binary crossover (SBX) as real-parameter re-combination operator w, crossover operator like uniform crossover (UC) it looks at one specific gene at a time , and single point crossover in a random crossover point , are used to test the suggested adaptive mutation operator technique (SI). Following that, 3 new NSGA-III algorithms (SBXAM, SIAM, as well as UCAM) were created. After then, such improved ones are used to handle a variety of Big-Opt challenges. The NSGA-III plus UC as well as adaptive NSGA-III UCAM which is mutation operator significantly outperforms the several NSGA-III algorithms, according to the findings.

In [14] proposed optimization based feature selection to enhance efficiency and reduce burden during procedure on mining also on model, the FS approach has been used for on-the-fly streaming data mining for large data. A Particle Swarm Optimization (PSO) a computational method, an evolutionary computation is Grammatical Evolution (GE), and then a Hybrid PSO-GE approach have been developed in this study for obtaining accuracy and having a least computational time for a query and for reducing the burden. The categorization approaches make it easier for users to get the information they need from a large database of transactions. Throughout years, database management solutions have been essential to businesses. As the quantity of data has grown exponentially, database aggregation has run into a conflict among privacy and efficiency. To preserve data privacy, all characteristics in conventional database aggregation have been encrypted. The results of the studies show that the hybrid PSO-GE approach is more efficient than other methods.

In [15] used Machine Learning classifier approach with a separate feature set, with feature selection (FS) being the primary mechanism of these feature-based classifiers. Firstly, the candidate feature set was found using the Minimum Redundancy – Maximum Relevance technique, which took into account the greatest connection with target class as well as minimal redundancy as in features (mRMR). Next, as key highlight feature set for the detection activities, the reliable feature set only with fewest features that may reach best performance is chosen. The suggested technique was put to the test on 2 data sets from Twitter's networking site, as well as

the results compared to those of other efficient approaches. In comparison to current approaches, the suggested classifier strategy performs better, according to the findings. Studies on 2 Twitter datasets revealed that the suggested strategy might produce more optimum and balancing performance than other current approaches. With a modest tweak in the feature set, the suggested technique may be used to identify phony accounts on a variety of social networks, which might be a future study's goal.

In [16] author presented MOFS-BDE, a novel multi-objective differential evolution technique for FS issues. Several additional operators have been created and included into MOFS-BDE, particularly probability focused binary mutation, 1-bit purifying search/purifying search. It show that MOFS-BDE is effective when work with FS difficulties by comparing it to four common FS techniques (Multi-objective Differential Evolution, Non-dominated Sorting Genetic Algorithm, Multi-Objective Particle Swarm Optimization, as well as multi-objective artificial bee colony) as well as a novel Multiobjective Evolutionary Algorithm MOEA/D method (MOEA/D-2TMFI). The paper's key conclusions may be summarized as follows. First suggested probability-focused binary mutation may significantly enhance the convergence of algorithm based on the optimum learning method. MOFS-BDE can provide suitable solutions for the majority of the test datasets, outperforming the 5 comparison techniques. Secondly unique one-bit purifying search fully exploits the features of feature selection issues. And last that one-step purifying selection intended to be a low refining procedure. Evolutionary feature selection techniques have higher computing difficulties than traditional approaches like the mRMR and sequential FS. The runtime time may increase the occurrence of samples/features increases. We'll also use meta-heuristic techniques to solve more difficult data mining issues like MOO multi-label and stream FS.

In [17] author proposed feature selection (FS) as a nondeterministic polynomial time problem task that seeks to choose specific and necessary/non-repititive feature subset for enhancing performance of classification model. Some heuristic approaches, like particle swarm optimization (PSO) is huge hit, but as the number of features grows, the solution space was enormous, subsequently reduces searchability. Recently discretization-based FS approaches translate feature domain search towards cut-point domain, considerably reducing solution space as well as improving performance. Proposing an evolved MOO model in discretization focused Feature Selection in this study, taking into account the conflicts between multiple goals. A flexible cut-point PSO (FCPSO) is presented in exploring essessential features and achieves the Pareto solutions. It may pick an undetermined value at cut-points to further process. To successfully discover significant aspects and delete the redundant data in FCPSO, particle updates as well as a unique adapting mutation operator are utilized alternately. Finally, a Pareto ensemble approach were devised producing many possible solutions focused upon

the Pareto set, then hierarchical solution used in procedure, in order to identify the optimal feature subset. This suggested model greatly exceeds existing techniques for test performance efficiency along the competitive amount to FS.

In [18] author presented a methodology for investigating a Pareto NS set using parallel multi-objective NSGA-II, becomes very popular, quick computation as well as high valuable optimization technique multi objective GA. Multi-label classification algorithms such as like most basic approaches Binary Relevance (BR), like ML method for problem transformation Classifier Chains (CC), or where Pruned Sets (PS) having concept of treating sets of labels as single labels., as well as Random k-Labelset extract and verify subsets of non-dominated data (RAkEL). During the classification phase of the methods, base classifiers like SVM supervised learning models like Support Vector Machines, Implementation J48-Decision Tree (J48), as well as A logistic model as Logistic Regression (LR) are used. Local Image descriptors recovered by 2 set of data, another popular MIR community from Flickr website with a sensor nodes Wireless Multimedia Sensor (WMS) produced from video recordings, are used in extensive experiments. Both MIR-Flickr and WMS datasets boost forecasting accuracy by 6.36 percent and 25.7 percent, respectively, although feature amount is greatly decreased. Here findings shows this methodology proposed by novel framework exceeds current best practices.

In [19]author discussed two types of optimization algorithms: single-objective (SOO) and multi-objective (MOO). MOO contains many objective problems to create the Pareto set, while single-objective optimization has only one objective function; hence, solving a multi-objective issue is difficult. This work introduces a novel multi-objective optimization approach (dubbed MWDEO) focused on an upgraded whale optimization algorithm (WOA) that combines the heuristic approach differential evolution (DE) with efficient stochastic opposition-based learning algorithms (OBL). The MWDEO employs the WOA for framework global exploration, the DE for exploitation of the search space, and the OBL in improving exploration as well as exploitation by creating the opposite values. A collection of 32 multi-objective test problems, as well as a set of CEC'2017 benchmark tasks, are used to assess the suggested approach. 9 state of the art MOO approaches are analyze for experimental outcomes. The findings indicated that the suggested MWDEO exceeded many other algorithms in the majority of the test cases, indicating it is successful and competitive in addressing various kinds of multi-objective issues.

III. MULTI-OBJECTIVE CLUSTERING BASED GWO (MOCGWO)

In the Multi-Objective clustering-based Grey Wolf Optimizer (MOCGWO), a secure-shaped external collection is combined with the GWO for redeemable and recovering the Pareto optimum explanations. This collection has been active to

describe the communal grading and pretend the shooting performance of older wolves in multi-objective pursuit places.

3.1 The data clustering problem

The clustering difficulties can be expressed as a multiple-impartial optimization problematic by concurrently enhancing N sum of impartial tasks.

Improved clustering:

$$P(A) = \text{Optimum} [p_1(A), p_2(A), \dots, p_n(A)] \quad (1)$$

Where, A is the dataset defining all possible set of clusters. P defines the dataset of impartial tasks which are essential for optimization .More precisely equation 1.can be rewrite as

Intended for an instance, if we consider all the purposes that we are commerce with lowest optimization. Then we can rewrite the Equation (1) as follows:

$$\text{Minimalize } P(A) = \min[p_1(A), p_2(A), \dots, p_n(A)] \quad (2)$$

Present day the Pareto optimality can be understood by concurrently reducing the N amount of impartial. A answer included in as a part of Pareto obverse only if it is conquered by no one.

For this following two keys can be taken for reflection T1 and T2 where $T_1, T_2 \in T$

The key T2 is supposed to be conquered by T1 if

$$p_i(T_1) \leq p_i(T_2), \forall i, i \in 1, 2, \dots, N \quad (3)$$

$$p_j(T_1) < p_j(T_2), \exists j, j \in 1, 2, \dots, N \quad (4)$$

Later, the Pareto optimum keys include of all non-conquered keys and provide the finest key from a set of those Pareto optimal keys which can be found on operator's requirement.

Multi-objective optimization (also known as multi-objective encoding, course optimization, multicriteria optimization, multi-attribute optimization or Pareto optimization) is an part of numerous standards decision making that is troubled with mathematical optimization glitches connecting additional than one impartial function to be enhanced concurrently. Multi-objective optimization has been practical in numerous arenas of discipline, counting engineering, finances and logistics where best choices essential to be occupied in the company of trade-offs amid two or more contradictory objects.

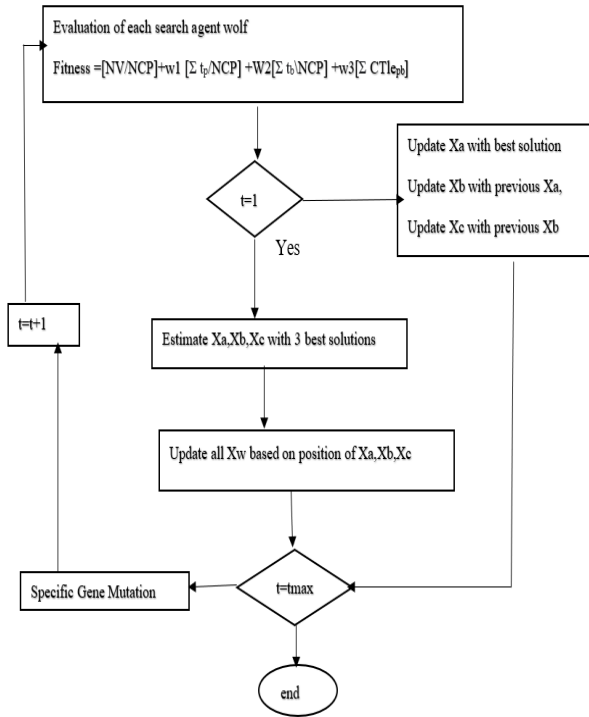


Fig 2. Multi-objective GWO Flowchart

3.2 Working of GWO

The simple working of MOGWO is inspired after GWO only but nearby are two individual two mechanisms in MOGWO. First module agrees to the documentation set and the other component is the main collection. Let an optimization problematic A essential to enhance three impartial purposes, explicitly p1, p2, and p3. Considering here can be two keys, G and H . A key G rules an extra key H , if G is restored by H in however one of the purposes pi and not inferior than other objects. If somehow, G does not control H and H does not control G, they are called uncontrolled keys. These keys are measured as better keys as compared to another keys in the residents.

To conserve the flow regulator in MOGWO an active controller is used that controls the actions of non-subject keys.

Subsequent rules of amount are followed on the store usual dataset:

1. Z can be added to ARC if a new solution Z is dominated by any other solution.
2. The dominated solution are deleted from ACH and Z is added if new solution Z dominated one or more solutions of ACH.
3. Z is added if both the above conditions fails.
4. The most crowded segments can be identified using grip mechanism and one of its solution is deleted if ACH is full. And solution Z is added to the least crowded segment to maintain the diversity.

5. If the ACH is full, the most crowded segments are identified using the grid mechanism and one of its solutions is deleted. Then, the solution W is added to the least crowded segment to maintain the diversity.

Consequently, MOGWO keeps the variety in the populace and reconnoiters individually section of the hunt interplanetary to discovery the best solutions. The location apprise reckonings and additional steps of MOGWO procedure conferring to GWO.

IV. RESULTS AND DISCUSSION

The proposed MOCGWO is implemented on 5 datasets. These datasets were taken from the UCI machine learning repository. The proposed model is simulated using MATLAB platform and shows the classification result with ensemble classifier. All the experiments are performed on Intel Core i5 CPU (3.70 GHz) and 8 GB of RAM and the main programming language is in MATLAB. The result was compared with other optimization technique such as PSO and GWO.

Analysis of the change in iteration numbers with fitness values is shown for the evaluation of sentiment analysis and accuracy. Result shows how the overall number of iterations and fitness values have changed over time. In Fig 3 Fitness value (FV) vs number of iteration were shown for PSO algorithm. The fitness decreases from 0.103 to 0.6 for 0-20 number of iteration and almost constant after 20 iterations. The accuracy of the PSO is evaluated on 5 datasets Brain Tumor, CNS, Lung, ionosphere and NSL_KDD. PSO have accuracy of 0.8333 for Brain Tumor, 0.975 for Lung, 0.975 for ionosphere, 0.97for NSL_KDD. Table 1 shows the comparison of Accuracy evaluation. Result shows that the proposed method outperformed all the methods in terms of accuracy. In Fig 4 FV vs no. of iteration were shown for GWO model. The fitness decreases from 0.12 to 0.03 for 0-20 number of iteration and almost constant after 20 iterations. The accuracy of the GWO is evaluated on 5 datasets Brain Tumor, CNS, Lung, ionosphere and NSL_KDD. PSO have accuracy of 0.944 for Brain Tumor, 1 FOR CNS, 0.975 for Lung, 0.957 for ionosphere, 0.98 for NSL_KDD. In Fig 5 FV vs no. of iteration were shown for MOCGWO model. The fitness decreases from 0-20 number of iteration and almost constant after 20 iterations. The accuracy of the MOCGWO is evaluated on 5 datasets Brain Tumor, CNS, Lung, ionosphere and NSL_KDD. MOCGWO have accuracy of 1 for Brain Tumor, 1 for CNS, 1 for Lung, 0.971 for ionosphere, 0.99 for NSL_KDD.

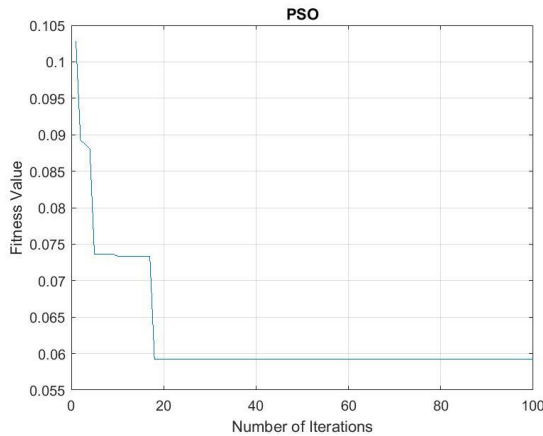


Fig. 3. FV vs Number of iteration for PSO Algorithm

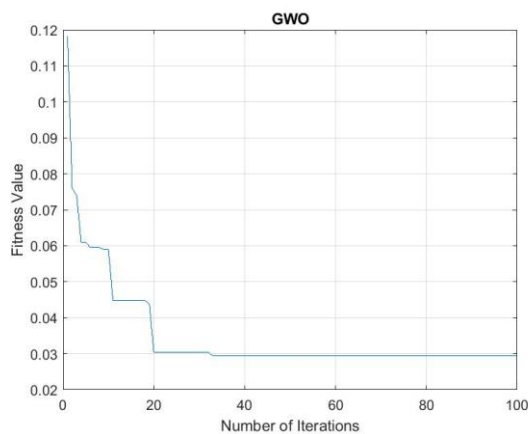


Fig. 4. FV vs No. of iteration for GWO Algorithm

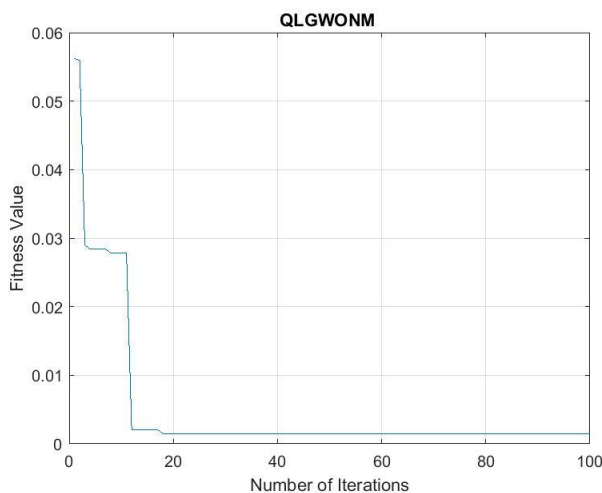


Fig. 5. Fitness value vs Number of iteration for MOCGWO Algorithm

Table 1: Comparison of Accuracy Evaluation

Dataset	PSO	GWO	MOCGWO
Brain Tumor	0.8333	0.944	1

CNS	1	1	1
Lung	0.975	0.975	1
Ionosphere	0.957	0.957	0.971
NSL_KDD	0.97	0.98	0.99

PSO= Particle Swarm Optimization, GWO= Gray Wolf Optimization, MOCGWO= Multi-objective Clustering Gray Wolf Optimization.

V. CONCLUSION

With the rapid development in technology, massive new digital technology implementations have engendered huge amounts of data at an extraordinary speed, including clip, picture, message, speech, and information derived from social relationships, as well as the increase of the Internet - of - things and edge computing. These data often have a high number of dimensions, which makes analysis of data and decision-making difficult. One of the biggest issues with Big Data is the enormous amount of characteristics or dimensions, which leads to the "curse of dimensionality" when using machine learning techniques, particularly classification algorithms. Feature selection is a useful strategy for improving learning performance by selecting tiny and informative feature subsets. Due to the wide and complicated search space, feature selection is a difficult process. Due to its flexibility and future global search ability, now several AI frameworks updated and have gotten huge interest from the FS community. For Big Data Analytics, this paper introduces a novel approach called MOCGWO based FS Algorithm. The result analysis was performed on different classification datasets and it was compared with other optimization methodology such as GWO and PSO. The shows efficacy of MOCGWO algorithm. Here, suggested work outperformed all current methodologies, according to the results. Finally, but always a chance of development, the multi-objective FS can develop and form a variety of hybrids approaches as well as adaptations based upon situation needs. Therefore, additional possible MOO research might employ the findings of this study to further investigate efficient strategies for solving contemporary issues in multi-objective FS.

References

- [1] Zhu, Li, Fei Richard Yu, Yige Wang, Bin Ning, and Tao Tang. "Big data analytics in intelligent transportation systems: A survey." IEEE Transactions on Intelligent Transportation Systems 20, no. 1 (2018): 383-398.
- [2] Batko, Kornelia, and Andrzej Ślęzak. "The use of Big Data Analytics in healthcare." Journal of big Data 9, no. 1 (2022): 1-24.
- [3] Mehta, Nishita, and Anil Pandit. "Concurrence of big data analytics and healthcare: A systematic review." International journal of medical informatics 114 (2018): 57-65.
- [4] Luo, Mi, Yifu Wang, Yunhong Xie, Lai Zhou, Jingjing Qiao, Siyu Qiu, and Yujun Sun. "Combination of feature selection and catboost for prediction: The first application to the estimation of aboveground biomass." Forests 12, no. 2 (2021): 216.
- [5] Too, Jingwei, and Seyedali Mirjalili. "A hyper learning binary dragonfly algorithm for feature selection: A COVID-19 case study." Knowledge-Based Systems 212 (2021): 106553.
- [6] Agrawal, Prachi, Hattan F. Abutarboush, Talari Ganesh, and Ali Wagdy Mohamed. "Metaheuristic algorithms on feature selection: A survey of

- one decade of research (2009-2019)." *IEEE Access* 9 (2021): 26766-26791.
- [7] C. Cimpanu, L. Ferariu, T. Dumitriu and F. Ungureanu, "Multi-Objective Optimization of Feature Selection procedure for EEG signals classification," 2017 E-Health and Bioengineering Conference (EHB), 2017, pp. 434-437, doi: 10.1109/EHB.2017.7995454.
- [8] Luo, Juanjuan, Dongqing Zhou, Lingling Jiang, and Huadong Ma. "A particle swarm optimization based multiobjective memetic algorithm for high-dimensional feature selection." *Memetic Computing* (2022): 1-17.
- [9] Hao, Jin-Kao; Legrand, Pierrick; Collet, Pierre; Monmarché, Nicolas; Lutton, Evelyne; Schoenauer, Marc (2012). [Lecture Notes in Computer Science] *Artificial Evolution Volume 7401 || A Rigorous Runtime Analysis for Quasi-Random Restarts and Decreasing Stepsize.* , 10.1007/978-3-642-35533-2(Chapter 4), 37–48. doi:10.1007/978-3-642-35533-2_4
- [10] BenSaid, Fatma, and Adel M. Alimi. "Online feature selection system for big data classification based on multi-objective automated negotiation." *Pattern Recognition* 110 (2021): 107629.
- [11] Guo, Jianmei, Jia Hui Liang, Kai Shi, Dingyu Yang, Jingsong Zhang, Krzysztof Czamecki, Vijay Ganesh, and Huiqun Yu. "SMTIBEA: a hybrid multi-objective optimization algorithm for configuring large constrained software product lines." *Software & Systems Modeling* 18, no. 2 (2019): 1447-1466.
- [12] Abdi, Yousef, and Mohammad-Reza Feizi-Derakhshi. "Hybrid multi-objective evolutionary algorithm based on search manager framework for big data optimization problems." *Applied Soft Computing* 87 (2020): 105991.
- [13] Yi, Jiao-Hong, Suash Deb, Junyu Dong, Amir H. Alavi, and Gai-Ge Wang. "An improved NSGA-III algorithm with adaptive mutation operator for Big Data optimization problems." *Future Generation Computer Systems* 88 (2018): 571-585.
- [14] Meera, S., and C. Sundar. "A hybrid metaheuristic approach for efficient feature selection methods in big data." *Journal of Ambient Intelligence and Humanized Computing* 12, no. 3 (2021): 3743-3751.
- [15] Rostami, Reza Ramzanzadeh, and Soheila Karbasi. "Detecting Fake Accounts on Twitter Social Network Using Multi-Objective Hybrid Feature Selection Approach." *Webology* 17, no. 1 (2020).
- [16] Zhang, Yong, Dun-wei Gong, Xiao-zhi Gao, Tian Tian, and Xiao-yan Sun. "Binary differential evolution with self-learning for multi-objective feature selection." *Information Sciences* 507 (2020): 67-85.
- [17] Zhou, Yu, Junhao Kang, Sam Kwong, Xu Wang, and Qingfu Zhang. "An evolutionary multi-objective optimization framework of discretization-based feature selection for classification." *Swarm and Evolutionary Computation* 60 (2021): 100770.
- [18] Karagoz, Gizem Nur, Adnan Yazici, Tansel Dokeroglu, and Ahmet Cosar. "A new framework of multi-objective evolutionary algorithms for feature selection and multi-label classification of video data." *International Journal of Machine Learning and Cybernetics* 12, no. 1 (2021): 53-71.
- [19] Ewees, Ahmed A., Mohamed Abd Elaziz, and Diego Oliva. "A new multi-objective optimization algorithm combined with opposition-based learning." *Expert Systems with Applications* 165 (2021): 113844.